-65-275

(INTERIM REPORT)

DATA FILE SIZE AND ITS RELATION TO THE BAYESIAN EFFECTIVENESS OF AN INFORMATION RETRIEVAL SYSTEM

Ugo O. Gagliardi

APRIL 1965

DECISION SCIENCES LABORATORY
ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L. G. Hanscom Field, Bedford, Massachusetts

Project 2806, Task 280609

(Prepared under Contract No. AF 19 (628)-5057 by Dunlap and Associates Incorporated, Darien, Connecticut.)

DDC AVAILABILITY NOTICE

Copies available from the Defense Documentation Center. (DDC)

DISSEMINATION NOTICE

DDC release to the Clearing House for Federal Scientific & Technical Information is authorized. (CFSTI) (Formerly OTS)

OTHER NOTICES

Do not return this copy. Retain or destroy.

ESD-TR 65-275

DATA FILE SIZE AND ITS RELATION TO THE BAYESIAN EFFECTIVENESS OF AN INFORMATION RETRIEVAL SYSTEM

Ugo O. Gagliardi

April 1965

FOREWORD

This study was accomplished under Project 2806, Task 280609, with Dunlap and Associates, Inc., Darien, Connecticut, Contract AF19(628)-5057. Contract Monitor is Dr. E. H. Shuford, Jr., ESRHT. Inclusive dates of research reported was 16 February to 15 April 1965, and it was submitted in April 1965. The author wishes to thank Dr. D. Promisel and Mr. R. J. Matteis for their assistance in the review and critique of this report. The author also wishes to acknowledge the combined support and encouragement he received from Dr. E. H. Shuford, Jr., Decision Sciences Laboratory.

This Technical Report has been reviewed and is approved.

Walter E. Organist
WALTER E. ORGANIST
Project Officer
Decision Sciences Laboratory

Roy Morgan
ROY MORGAN, Colonel, USAF
Director
Decision Sciences Laboratory

ABSTRACT

A simple Bayesian measure of system effectiveness for information retrieval systems is proposed. The measure combines the recall and precision ratios of an information system with the utility structure of the system user. Using the measure, it is possible to show that effective systems are possible only under a very narrow set of conditions. In particular, it is shown that using present state-of-the-art indexing, it is not possible to have effective systems with file sizes much in excess of 100, 000 documents.

TABLE OF CONTENTS

ILLUSTRATIONS

SECTION I

INTRODUCTION

The development in the last two decades of several new technologies for storing and retrieving information has resulted in a growing interest in the design and development of large scale information retrieval systems. Following an initial period of enthusiasm about the potential of the new technologies, a more sobering attitude has developed due to the realization that many of the existing and proposed systems are not justified on a cost-effectiveness basis. In this report we wish to present a model for evaluating the effectiveness of information retrieval systems which is based on Bayesian statistical theory. This model relates the effectiveness of a retrieval system to the statistical characteristics of the retrieval process and the value structure of the requestor. The model can and does identify the operational conditions under which, using present state-of-the-art techniques, one cannot expect to obtain sufficient system performance to justify the development of a system. An information retrieval system is basically a system composed of a file of documents and procedures for partitioning it under the control of the requestor. The collection of documents is mapped into a suitable multi-dimensional space by the indexing process. Indexing results in the association of one or more tags (index terms) to document of the file. The index space is usually a Boolean lattice, whose independent variables are the index terms. The index space is partitioned by defining a truth function over the index space.

The input to the information retrieval system is a statement describing the class of concepts on which one desires additional information. The output of the system is a partition of the documents into two clases: the retrieved and non-retrieved documents. Then by definition, all the devices, operators, and processes, which are utilized in transforming the input request into the output, are elements of the information retrieval system.

The approach taken in this report is to characterize the overall error process by two parameters: the recall and the precision ratio (Ref. 4, 5). These ratios give, respectively, an indication of the average degree of completeness and the average degree of purity (i. e., lack of irrelevant material) of the search. These two parameters are then used to express a complete model of the error process of the retrieval system in a form which is compatible with the formal evaluation technique developed in References 1 and 2. No attempt is made at this time to analyze the error process in its components: indexing error process, request coding error process, and search error process.

SECTION II

A BAYESIAN EFFECTIVENESS MODEL FOR RETRIEVAL SYSTEMS

In References 1 and 2 a scalar measure of effectiveness for information systems is derived. This derivation is also given in Appendix B. The effectiveness measure represents the gain in expected utility which results when a rational decision maker utilizes the outputs of the information system. The effectiveness measure is given by the following formula:

$$\epsilon(Q) = [UD_\pi Q]^* \xi - [U\pi]^* \tag{1}$$

where

Q	=	$\{P(Y^i \mid X^j)\}$	The Information System Model.
U	=	$\{U(A^k \mid X^i)\}$	The utility structure of the decision maker.
π	=	$\{P(X^i)\}$	The prior distribution.
D_π	=	Diagonal form of π.	
ξ	=	Vector of all ones.	
$[\ ]^*$	=	Operation which takes the largest component of the column(s) inside.	

A simple but interesting model for the evaluation of a retrieval system using this effectiveness measure is obtained as follows:

Let us introduce (Refs. 3, 4, 6, 7) the recall ratio and the precision ratio of the retrieval system. By precision ratio one means the ratio of relevant material labeled retrievable to the material labeled retrievable. Thus, this ratio is equal to

$$p = \frac{NP_r\,[\text{Item is relevant and is labeled retrievable}]}{NP_r\,[\text{Item is labeled retrievable}]}$$

$$= \frac{P(\rho, R)}{P(\rho)} = P(R \mid \rho) \tag{2}$$

where

ρ = item being examined is labeled retrievable.

R = item being examined is relevant

By recall ratio we mean the ratio of relevant material which is labeled retrievable to the total amount of material which is relevant. Thus,

$$r = \frac{NP_r\text{ [Item is relevant and is labeled retrievable]}}{NP_r\text{ [Item is relevant]}}$$

$$= \frac{P(\rho, R)}{P(R)} = P(\rho|R) \tag{3}$$

But by Bayes rule,

$$p = P(R|\rho) = \frac{P(\rho|R)P(R)}{P(\rho, R) + P(\rho, \bar{R})} = \frac{P(\rho|R)P(R)}{P(\rho|R)P(R) + P(\rho|\bar{R})P(\bar{R})}$$

and if we let

$q = P(R)$

$1-q = P(\bar{R})$

we obtain from (2) and (3) that

$$P(\rho|\bar{R}) = r\frac{q}{1-q}\;\frac{1-p}{p} = r\beta$$

$$P(\rho|R) = r$$

which together with

$$P(\bar{\rho}|R) = 1-P(\rho|R) = 1-r$$

$$P(\bar{\rho}|\bar{R}) = 1-P(\rho|\bar{R}) = 1-r\beta \tag{4}$$

completely determine a simple binary model of the retrieval system to be:

$$Q = \begin{vmatrix} r & 1-r \\ r\beta & 1-r\beta \end{vmatrix}$$

where

r = Recall ratio of the information retrieval system.

$$\beta = \frac{q}{1-q} \quad \frac{1-p}{p}$$

p = Precision ratio of the information retrieval system.

q = Probability that an item in the collection will be relevant to the specific query.

The information system output is to be used to decide which of the two actions retrieve or do not retrieve should be taken. Using this simple model, the retrieval problem can be considered a two-state, two-act decision problem. The two states[1] are "the item is relevant to the inquiry," "the item is not relevant to the inquiry." Thus the utility structure for such a decision task is:

---

[1] Let us briefly digress on the validity of the two-state decision theoretic formulation of the retrieval problem. The point, of course, is whether a two-valued view of relevancy is appropriate; Reference 7 attacks the two-valued view of relevancy as incorrect, but the decision theoretical formulation of the problem shows that this is an unnecessarily extreme position since the substitution of a two-valued relevancy for a multivalued relevancy is tantamount to coalescing the states into two states with the consequent averaging of the corresponding entries of the utility structure. Thus, the adoption of a two-valued relevancy will reduce the accuracy in computing the effectiveness, but the errors so committed are going to be bounded (and reasonable, in general) in light of the properties of convex combinations. It seems to us that it is useless to advocate a multivalued view of relevancy in view of the difficulty and inaccuracy inherent in the measurement of relevancy since one would be achieving, in most cases, only a pseudo higher accuracy.

$$\begin{array}{lcc} & R & \bar{R} \\ \text{Retrieve} & U_{11} & U_{12} \\ \text{Do Not Retrieve} & U_{21} & U_{22} \end{array}$$

The utility structure above gives a value to each of the following contingencies:

	Document Is Relevant	Document Is Not Relevant
Retrieve Document	Hit	False Drop
Do Not Retrieve Document	Miss	Correct Rejection

Thus for example, U_{12} is the value of a false drop. Let us now consider the following utility structure:

$$U = \begin{vmatrix} \gamma & -1 \\ -\alpha & 0 \end{vmatrix} \tag{6}$$

Such a structure is of absolute generality since it can be shown that the addition of a constant to each entry of the utility structure does not change the value of the effectiveness function given in (1); thus one can always zero one of the entries of the utility structure. Also the division of all the entries of the utility structure by the same constant divides the effectiveness measure by the same constant.[2] Thus setting of one entry to ± 1 is tantamount to the selection of the unit of measure of effectiveness. Thus, substituting in (1) one obtains

$$\epsilon(Q) = \left[\begin{vmatrix} \gamma & -1 \\ -\alpha & 0 \end{vmatrix} \begin{vmatrix} q & 0 \\ 0 & 1-q \end{vmatrix} \begin{vmatrix} r & 1-r \\ r\beta & 1-r\beta \end{vmatrix} \right]^{*} \xi + - \left[\begin{vmatrix} \gamma & -1 \\ -\alpha & 0 \end{vmatrix} \begin{vmatrix} q \\ 1-q \end{vmatrix} \right]^{*} \tag{7}$$

---

[2] The effectiveness function is a piece-wise linear function of the utility entries.

which, after a little algebra, (7) becomes

$$\epsilon(Q) = \begin{vmatrix} r(q\gamma-\beta)(1-q) & q\gamma(1-r)-(1-q)(1-r\beta) \\ -\alpha qr & -\alpha q(1-r) \end{vmatrix}^{*} \xi + - \begin{vmatrix} \gamma q-(1-q) \\ -\alpha q \end{vmatrix}^{*} \quad (8)$$

The above (8) represents thus the simplest evaluation model for retrieval systems which is consistent with a Bayesian statistical viewpoint as discussed in Reference 4. The above model has the following five dimensions:

r = The recall ratio.

p = The precision ratio. (in $\beta = \frac{q}{1-q} \quad \frac{1-p}{p}$)

q = The density of relevant material.

α = The loss ratio of misses vs. false drops.

γ = The utility ratio of hits vs. false drops.

The above model was coded as a FORTRAN II procedure and then compiled on an SDS 920 computer. The version of this program which was actually used in calculating the data reported in this report is listed in Appendix I.

Using the program of Appendix I, three computational experiments were executed. These experiments show that the behavior of the model (8) is determined predominantly by two parameters:

$$\theta = \alpha q \quad \text{and} \quad \lambda = \gamma q$$

In experiment A, which has the design illustrated in Figure 1, the parameter θ is much greater than one, while the parameter λ spans the range 10^{-5} to 10^{6}. As it can be seen from the results, the effectiveness of the ISR (Information Storage and Retrieval System) is uniformly zero within the design ranges of $r \leq .95$ and $p \leq .95$. This result indicates that unless one is prepared to furnish a system with a recall ratio of better than 95% and a precision ratio of better than 50%[3], one is as well off with no information retrieval system at all.

[3] This figure is obtained from results not included in Appendix I.

$\theta = 10^2$

	$\gamma = 0$	$\gamma = 10$	$\gamma = 10^3$	$\gamma = 10^5$	$\gamma = 10^7$
$q = 10^{-1}$					
$q = 10^{-3}$					
$q = 10^{-5}$					
$q = 10^{-6}$					

Figure 1. Experiment A.

Reference 6, quoting Cleverdon, states "indications are that information retrieval systems are generally operating at a recall ratio of 70 to 90 per cent with relevance (i. e., precision ratio) in the range of 8 to 20 per cent." Thus, it seems that state-of-the-art techniques are quite far from allowing, economically, a design level of (> 50%, > 95%).[4]

In experiment B, which has the design illustrated in Figure 2, the value of θ is unity and λ ranges from 10^{-5} to 10^{6}. It is apparent from the results of this experiment that if $\lambda > 1$, the effectiveness of the ISR becomes uniformly null for all but the highest values of the recall ratio. The behavior of the effectiveness function is quite similar in this case to the previous experiment. We can thus conclude that if $\theta >> 1$ OR $\lambda >> 1$, the effectiveness of all ISR designs except those with extremely high values of recall and precision ratios is null.

Experiment C has the design illustrated in Figure 3. In this experiment $\theta = 10^{-3}$, i. e., $\theta << 1$ while λ ranges from 10^{-8} to 10^{4}. It is seen that the design divides into three regions: the region $\lambda >> 1$ for which the effectiveness is predominantly zero; the region $\lambda \equiv 1$ for which the effectiveness is predominantly non-zero; and the region $\lambda << 1$ for which the effectiveness is again predominantly zero. The first region is the result of the previously stated property that if $\theta >> 1$ or $\lambda >> 1$, the effectiveness will be predominantly null. The third region is evidence that if $\theta << 1$ <u>and</u> $\lambda << 1$, the effectiveness will be predominantly null.

We can thus conclude that the effectiveness of an ISR is null for all but extremely demanding levels of design if <u>either</u>

$$\theta >> 1 \text{ or } \lambda >> 1$$

<u>or</u>

$$\theta << 1 \text{ and } \lambda << 1$$

where >> and << mean different by more than one full order of magnitude.

---

[4] Design levels will be indicated with the order pair (p, r).

$\theta = 1$

$q = 10^{-1}$					
$q = 10^{-2}$					
$q = 10^{-3}$					
$q = 10^{-5}$					
$q = 10^{-7}$					
	$\gamma = 10$	$\gamma = 10^2$	$\gamma = 10^3$	$\gamma = 10^5$	$\gamma = 10^7$

Figure 2. Experiment B.

$\theta = 10^{-3}$

$q = 10^{-2}$					
$q = 10^{-3}$					
$q = 10^{-5}$					
$q = 10^{-6}$					
$q = 10^{-7}$					
$q = 10^{-8}$					
	$\gamma = 0$	$\gamma = 1$	$\gamma = 10^2$	$\gamma = 10^4$	$\gamma = 10^6$

Figure 3. Experiment C.

Limits on the Acceptable Density of Relevant Material

Both the parameters θ and λ contain the factor q which is the probability that an item in the collection is relevant to a specific query. The results obtained can be used to establish lower and upper bounds for the value of q.

If one assumes that the state of the art of information retrieval consents to obtain the levels of performance[5]

$$\begin{aligned} r &= .8 \\ p &= .2 \end{aligned}$$

then our result indicates that in order to get a non-negligible effectiveness, one must have

$$\theta \leq 1 \text{ and } \lambda \leq 1$$

as well as

$$\frac{1}{10} \leq \theta \text{ or } \frac{1}{10} \leq \lambda$$

The first condition is equivalent to

$$q \leq \frac{1}{\alpha} \text{ and } q \leq \frac{1}{\gamma}$$

which in turn can be expressed as

$$q \leq \begin{vmatrix} \dfrac{1}{\alpha} \\ \dfrac{1}{\gamma} \end{vmatrix}^{+}$$

where the operator + takes the smallest component of the vector. The second condition is equivalent to

$$q \geq \frac{1}{10\alpha} \text{ or } q \geq \frac{1}{10\gamma}$$

---

[5] This assumption is extremely optimistic since according to the only empirical evidence available to date (Ref. 3), one cannot attain both of these levels simultaneously with the present state of the art.

which is equivalent to

$$q \geq \begin{vmatrix} \dfrac{1}{10\alpha} \\ \\ \dfrac{1}{10\gamma} \end{vmatrix}^{+}$$

Thus

$$\begin{vmatrix} \dfrac{1}{10\alpha} \\ \\ \dfrac{1}{10\gamma} \end{vmatrix}^{+} \leq q \leq \begin{vmatrix} \dfrac{1}{\alpha} \\ \\ \dfrac{1}{\gamma} \end{vmatrix}^{+} \qquad (9)$$

If

$$\delta = \begin{vmatrix} \alpha \\ \\ \gamma \end{vmatrix}^{*}$$

then (9) can be written as

$$\frac{1}{10\delta} \leq q \leq \frac{1}{\delta} \qquad (10)$$

Systems which fall above the upper bound correspond to files which are interrogated just as effectively (when compared to present day indexing and retrieval techniques) by exhaustive examination of all the items in the file, i. e., by dispensing with an ISR. Systems which fall below the lower bound correspond to systems which are never utilized since the elimination of the non-relevant material which is retrieved requires an effort that is greater than the value of the relevant material retrieved. In recent years, since the computer age started, many such retrieval Leviathans have been developed and the common experience with them agrees with what this simple Bayesian theory predicts, namely, that they are not being used.

It is well to point out here the critical need for research directed to developing superior indexing and retrieval techniques since it is only through the availability of such superior techniques that the size "range" will be broadened.

SECTION III

CONCLUSIONS

We have shown that assuming very optimistic levels of performance for the present state of the art, one has only a very narrow range for the density of relevant material conducive to feasible system designs. The density of relevant material can be expressed as the ratio $N_R | N$ where N_R is the number of relevant items and N the total number of items in the collection. This density will vary greatly with the individual question. We can obtain an indication of what are the probable feasible collection sizes by fixing a value of N_R which is "reasonable." Here, of course, we are interested in order of magnitudes. Clearly, $N_R = 10^2$ is the largest value that can be considered reasonable.

In the above formulations, the number α represents the number of irrelevant documents which a user is willing to examine to avoid missing a relevant document. On the other hand, the number γ represents the number of irrelevant documents which a user is willing to examine to find a relevant document. Thus, $\delta = \left|\frac{\alpha}{\gamma}\right|^*$ represents the greatest number of irrelevant documents the user is willing to examine for one positive outcome. The value of δ will depend on the amount of work that the examination of one document entails. Thus, δ can be increased by techniques of summarization and abstracting. It seems reasonable, though, to assume that δ will seldom exceed 10^2 and, thus, $N_R \delta$ will seldom exceed 10^4 with the result that the size of files which are amenable to effective retrieval should at best be in the range $10^4 \leq N \leq 10^5$.

This conclusion implies that collections of documents in the multimillion range cannot be effectively interrogated using a single step retrieval process. In other words, this file has to be subdivided into many files each within the range indicated by equation (10) and the appropriate file should be selected by a preselection stage.

REFERENCES

1 Gagliardi, U. O. "Mathematical Programming Techniques for Information Systems Design." ESD Technical Documentary Rpt. No. ESD-TDR-64-530, July 1964.

2. Gagliardi, U. O. "A Bayesian viewpoint in the design of information systems." Invited paper to be published in the Journal of the Japanese Operations Research Society.

3. Cleverdon, C. W. "The Aslib Cranfield Research Project on the Comparative Efficiency of Indexing Systems." Aslib Proceedings, Vol. 12, No. 12, Dec. 1960.

4. Cleverdon, C. W. "Report on the Testing and Analysis of an Investigation into the Efficiency of Indexing Systems." College of Aeronautics, Cranfield, England, October 1962.

5. Raiffa, H. and Schlaifer, R. "Applied Statistical Decision Theory." Harvard University, 1961.

6. "Centralization and Documentation." Arthur D. Little, Inc., 1964.

7. A. D. Little, Inc., "Automatic Message Retrieval." ESD Technical Documentary Report No. ESD-TDR-63-673, November 1963.

APPENDIX I

THE COMPUTER PROGRAM AND RESULTS OF EXPERIMENTS A, B, C

```
 C     BAYESIAN EFFECTIVENESS OF RETRIEVAL SYSTEM 22 MARCH 65 UOG
 C     NON LINEAR AXIS
       DIMENSION C[2,2],B[2],RECALL[5],PRECIS[5],TOP[2],EFF[5,5]
       PAUSE
 9     READ 1,Q,ALPHA,GAMMA
 1     FORMAT [3[E15.5]]
       B[1]=GAMMA*Q-[1.0-Q]
       B[2]=-ALPHA*Q
       TOPB=B[1]
       IF[B[1]-B[2]]6,7,7
  6    TOPB =B[2]
  7    DO 5 I=1,5
       DO 5 J=1,5
       RECALL[I]=1.0-.025*2**I
       PRECIS[J]=1.0-.025*2**J
       BETA=[Q/[1.0-Q]]*[[1.0-PRECIS[J]]/PRECIS[J]]
       C[1,1]=RECALL[I]*[Q*GAMMA-BETA*[1.0-Q]]
       C[1,2]=Q*GAMMA*[1.0-RECALL[I]]-[1.0-Q]*[1.0-RECALL[I]*BETA]
       C[2,1]=-ALPHA*RECALL[I]*Q
       C[2,2]=-ALPHA*[1.0-RECALL[I]]*Q
       TOP[1]=C[1,1]
       IF[C[1,1]-C[2,1]]2,3,3
  2    TOP[1]=C[2,1]
  3    TOP[2]=C[1,2]
       IF[C[1,2]-C[2,2]]4,5,5
  4    TOP[2]=C[2,2]
       IF [SENSE SWITCH 1] 11,5
 11    TYPE 10,I,J,C,B,TOP,TOPB
 10    FORMAT [$I=$,I2,5X,$J=$,I2//$C$//4[F9.4]//$B$//2[F9.4]//
      1$TOP$//2[F9.4]//$TOPB=$,F9.4]
  5    EFF[I,J]=TOP[1]+TOP[2]-TOPB
       TYPE 8,Q,ALPHA,GAMMA,[[EFF[I,J],I=5,1,-1], J=1,5]
 8     FORMAT[$Q=$,E15.5,3X,$ALPHA=$,E15.5,3X,$GAMMA=$,E15.5/////
      1$EFFECTIVENESS TABLE $///[5 F8.2/]]
  12   FORMAT [/////////////////////////////////////////////////]
       TYPE 12
       GO TO 9
       END
```

PROGRAM ALLOCATION

00006 C	00016 B	00022 RECALL	00034 PRECIS
00046 TOP	00052 EFF	00134 I	00135 J
00136 Q	00140 ALPHA	00142 GAMMA	00144 TOPB
00146 BETA			

EXPERIMENT A

$\theta = 10^2$

$q = 10^{-1}$	Null at all Sample Points	Null at all Sample Points	Null at all Sample Points	Null at all Sample Points	Null at all Sample Points
$q = 10^{-3}$	Null at all Sample Points	Null at all Sample Points	Null at all Sample Points	Null at all Sample Points	Null at all Sample Points
$q = 10^{-5}$	Null at all Sample Points	Null at all Sample Points	Null at all Sample Points	Null at all Sample Points	Null at all Sample Points
$q = 10^{-6}$	Null at all Sample Points	Null at all Sample Points	Null at all Sample Points	Null at all Sample Points	Null at all Sample Points
	$\gamma = 0$	$\gamma = 10$	$\gamma = 10^3$	$\gamma = 10^5$	$\gamma = 10^7$

The sample points are the Carthesian product of the two following sets:

$p = .2,\ .6,\ .8,\ .9,\ .95$
$r = .2,\ .6,\ .8,\ .9,\ .95$

EXPERIMENT B

$\theta = 1$

$q = 10^{-1}$	Mostly Not Null	Not Null, Only in the Upper Right Corner	Null at All Sample Points	Null at All Sample Points	Null at All Sample Points
$q = 10^{-2}$	Not Null	Mostly Not Null	Not Null, Only in the Upper Right Corner	Null at All Sample Points	Null at All Sample Points
$q = 10^{-3}$	Not Null	Not Null	Mostly Not Null	Null at All Sample Points	Null at All Sample Points
$q = 10^{-5}$	Not Null	Not Null	Not Null	Mostly Not Null	Null at All Sample Points
$q = 10^{-7}$	Not Null	Not Null	Not Null	Not Null	Mostly Not Null
	$\gamma = 10$	$\gamma = 10^2$	$\gamma = 10^3$	$\gamma = 10^5$	$\gamma = 10^7$

The following pages contain the actual printout of the results. The rows of the result matrix correspond to distinct values of p and the columns to distinct values of r. Both variables vary over the set

.2, .6, .8, .9, .95

Q= 0.99999E-01 ALPHA= 0.10000E 02 GAMMA= 0.10000E 02

EFFECTIVENESS TABLE

-0.00	0.10	0.50	0.70	0.80
-0.00	0.09	0.49	0.69	0.79
-0.00	0.08	0.48	0.68	0.78
-0.00	0.06	0.45	0.64	0.74
-0.00	-0.00	0.18	0.34	0.42

Q= 0.99999E-01 ALPHA= 0.10000E 02 GAMMA= 0.10000E 03

EFFECTIVENESS TABLE

-0.00	-0.00	-0.00	-0.00	0.34
-0.00	-0.00	-0.00	-0.00	0.34
-0.00	-0.00	-0.00	-0.00	0.33
-0.00	-0.00	-0.00	-0.00	0.29
-0.00	-0.00	-0.00	-0.00	-0.00

Q= 0.99999E-01 ALPHA= 0.10000E 02 GAMMA= 0.10000E 04

EFFECTIVENESS TABLE

-0.00	-0.00	-0.00	-0.00	0.00
-0.00	-0.00	0.00	-0.00	-0.00
-0.00	-0.00	-0.00	-0.00	-0.00
-0.00	-0.00	-0.00	0.00	-0.00
-0.00	-0.00	-0.00	-0.00	-0.00

Q= 0.99999E-01 ALPHA= 0.10000E 02 GAMMA= 0.10000E 06

EFFECTIVENESS TABLE

-0.00	-0.00	-0.00	-0.00	-0.00
-0.00	-0.00	0.00	-0.00	-0.00
0.00	-0.00	-0.00	-0.00	-0.00
-0.00	-0.00	-0.00	-0.00	-0.00
-0.00	-0.00	-0.00	-0.00	-0.00

Q= 0.99999E-01 ALPHA= 0.10000E 02 GAMMA= 0.10000E 08

EFFECTIVENESS TABLE

-0.00	-0.00	-0.00	-0.00	-0.00
-0.00	-0.00	-0.00	-0.00	-0.00
-0.00	-0.00	-0.00	-0.00	-0.00
-0.00	-0.00	-0.00	-0.00	-0.00
-0.00	-0.00	-0.00	-0.00	-0.00

Q=   0.99999E-02   ALPHA=   0.10000E 03   GAMMA=   0.10000E 02

EFFECTIVENESS TABLE

0.11	0.55	0.77	0.88	0.93
0.11	0.55	0.77	0.88	0.93
0.11	0.55	0.77	0.88	0.93
0.11	0.55	0.76	0.87	0.93
0.10	0.53	0.74	0.84	0.90

Q= 0.99999E-02 ALPHA= 0.10000E 03 GAMMA= 0.10000E 03

EFFECTIVENESS TABLE

-0.00	0.19	0.59	0.79	0.89
-0.00	0.19	0.59	0.79	0.89
-0.00	0.19	0.59	0.79	0.89
-0.00	0.19	0.58	0.78	0.88
-0.00	0.17	0.56	0.75	0.85

Q= 0.99999E-02 ALPHA= 0.10000E 03 GAMMA= 0.10000E 04

EFFECTIVENESS TABLE

-0.00	-0.00	-0.00	-0.00	0.44
-0.00	-0.00	0.00	-0.00	0.44
-0.00	-0.00	0.00	-0.00	0.44
-0.00	-0.00	-0.00	-0.00	0.43
-0.00	-0.00	-0.00	-0.00	0.40

Q= 0.99999E-02 ALPHA= 0.10000E 03 GAMMA= 0.10000E 06

EFFECTIVENESS TABLE

-0.00	-0.00	0.00	-0.00	0.00
-0.00	-0.00	-0.00	-0.00	-0.00
-0.00	-0.00	-0.00	-0.00	-0.00
-0.00	-0.00	-0.00	-0.00	-0.00
-0.00	-0.00	-0.00	-0.00	-0.00

Q= 0.99999E-02 ALPHA= 0.10000E 03 GAMMA= 0.10000E 08

EFFECTIVENESS TABLE

-0.00	-0.00	-0.00	-0.00	0.00
-0.00	-0.00	-0.00	-0.00	-0.00
-0.00	-0.00	-0.00	-0.00	0.00
-0.00	-0.00	-0.00	-0.00	-0.00
-0.00	-0.00	-0.00	-0.00	-0.00

Q= 0.99999E-03 ALPHA= 0.10000E 04 GAMMA= 0.10000E 02

EFFECTIVENESS TABLE

0.19	0.59	0.80	0.90	0.95
0.19	0.59	0.80	0.90	0.95
0.19	0.59	0.80	0.90	0.95
0.19	0.59	0.80	0.90	0.95
0.19	0.59	0.79	0.89	0.94

Q= 0.99999E-03 ALPHA= 0.10000E 04 GAMMA= 0.10000E 03

EFFECTIVENESS TABLE

0.12	0.56	0.78	0.89	0.94
0.12	0.56	0.78	0.89	0.94
0.12	0.56	0.78	0.89	0.94
0.12	0.56	0.78	0.89	0.94
0.12	0.56	0.78	0.89	0.94

Q= 0.99999E-03 ALPHA= 0.10000E 04 GAMMA= 0.10000E 04

EFFECTIVENESS TABLE

-0.00	0.20	0.60	0.80	0.90
-0.00	0.20	0.60	0.80	0.90
-0.00	0.20	0.60	0.80	0.90
-0.00	0.20	0.60	0.80	0.90
-0.00	0.20	0.60	0.80	0.90

Q= 0.99999E-03 ALPHA= 0.10000E 04 GAMMA= 0.10000E 06

EFFECTIVENESS TABLE

-0.00	-0.00	-0.00	-0.00	-0.00
-0.00	-0.00	-0.00	-0.00	-0.00
-0.00	-0.00	-0.00	-0.00	-0.00
-0.00	-0.00	-0.00	-0.00	-0.00
-0.00	-0.00	-0.00	-0.00	-0.00

Q= 0.99999E-03 ALPHA= 0.10000E 04 GAMMA= 0.10000E 08

EFFECTIVENESS TABLE

-0.00	-0.00	-0.00	-0.00	-0.00
-0.00	-0.00	-0.00	-0.00	-0.00
-0.00	-0.00	-0.00	-0.00	-0.00
-0.00	-0.00	-0.00	-0.00	-0.00
-0.00	-0.00	-0.00	-0.00	-0.00

Q= 0.99999E-05 ALPHA= 0.10000E 06 GAMMA= 0.10000E 02

EFFECTIVENESS TABLE

0.20	0.60	0.80	0.90	0.95
0.20	0.60	0.80	0.90	0.95
0.20	0.60	0.80	0.90	0.95
0.20	0.60	0.80	0.90	0.95
0.20	0.60	0.80	0.90	0.95

Q= 0.99999E-05 ALPHA= 0.10000E 06 GAMMA= 0.10000E 03

EFFECTIVENESS TABLE

0.20	0.60	0.80	0.90	0.95
0.20	0.60	0.80	0.90	0.95
0.20	0.60	0.80	0.90	0.95
0.20	0.60	0.80	0.90	0.95
0.20	0.60	0.80	0.90	0.95

Q= 0.99999E-05 ALPHA= 0.10000E 06 GAMMA= 0.10000E 04

EFFECTIVENESS TABLE

0.19	0.60	0.80	0.90	0.95
0.19	0.60	0.80	0.90	0.95
0.19	0.60	0.80	0.90	0.95
0.19	0.60	0.80	0.90	0.95
0.19	0.60	0.80	0.90	0.95

Q=   0.99999E-05   ALPHA=   0.10000E 06   GAMMA=   0.10000E 06

EFFECTIVENESS TABLE

-0.00	0.20	0.60	0.80	0.90
-0.00	0.20	0.60	0.80	0.90
-0.00	0.20	0.60	0.80	0.90
-0.00	0.20	0.60	0.80	0.90
-0.00	0.20	0.60	0.80	0.90

Q= 0.99999E-05 ALPHA= 0.10000E 06 GAMMA= 0.10000E 08

EFFECTIVENESS TABLE

-0.00	-0.00	-0.00	-0.00	-0.00
-0.00	-0.00	-0.00	-0.00	-0.00
-0.00	-0.00	-0.00	0.00	0.00
-0.00	-0.00	-0.00	0.00	-0.00
-0.00	-0.00	-0.00	-0.00	-0.00

Q= 0.99999E-07 ALPHA= 0.10000E 08 GAMMA= 0.10000E 02

EFFECTIVENESS TABLE

0.20	0.60	0.80	0.90	0.95
0.20	0.60	0.80	0.90	0.95
0.20	0.60	0.80	0.90	0.95
0.20	0.60	0.80	0.90	0.95
0.20	0.60	0.80	0.90	0.95

Q= 0.99999E-07 ALPHA= 0.10000E 08 GAMMA= 0.10000E 03

EFFECTIVENESS TABLE

0.20	0.60	0.80	0.90	0.95
0.20	0.60	0.80	0.90	0.95
0.20	0.60	0.80	0.90	0.95
0.20	0.60	0.80	0.90	0.95
0.20	0.60	0.80	0.90	0.95

Q= 0.99999E-07 ALPHA= 0.10000E 08 GAMMA= 0.10000E 04

EFFECTIVENESS TABLE

0.20	0.60	0.80	0.90	0.95
0.20	0.60	0.80	0.90	0.95
0.20	0.60	0.80	0.90	0.95
0.20	0.60	0.80	0.90	0.95
0.20	0.60	0.80	0.90	0.95

Q= 0.99999E-07 ALPHA= 0.10000E 08 GAMMA= 0.10000E 06

EFFECTIVENESS TABLE

0.19	0.60	0.80	0.90	0.95
0.19	0.60	0.80	0.90	0.95
0.19	0.60	0.80	0.90	0.95
0.19	0.60	0.80	0.90	0.95
0.19	0.60	0.80	0.90	0.95

Q= 0.99999E-07 ALPHA= 0.10000E 08 GAMMA= 0.10000E 08

EFFECTIVENESS TABLE

-0.00	0.20	0.60	0.80	0.90
-0.00	0.20	0.60	0.80	0.90
-0.00	0.20	0.60	0.80	0.90
-0.00	0.20	0.60	0.80	0.90
-0.00	0.20	0.60	0.80	0.90

EXPERIMENT C

$\theta = 10^{-3}$

$q = 10^{-2}$	Null at All Sample Points	Non Null in Upper Right Quadrant	Not Null	Null at All Sample Points	Null at All Sample Points
$q = 10^{-3}$	Null at All Sample Points	Null at All Sample Points	Not Null	Now Null for All r = .95 Points	Null at All Sample Points
$q = 10^{-5}$	Null at All Sample Points	Null at All Sample Points	Null at All Sample Points	Not Null	Now Null for All r = .95 Points
$q = 10^{-6}$	Null at All Sample Points	Null at All Sample Points	Null at All Sample Points	Mostly Not Null	Not Null
$q = 10^{-7}$	Null at All Sample Points	Null at All Sample Points	Null at All Sample Points	Null at All Sample Points	Not Null
$q = 10^{-8}$	Null at All Sample Points	Null at All Sample Points	Null at All Sample Points	Null at All Sample Points	Mostly Not Null
	$\gamma = 0$	$\gamma = 1$	$\gamma = 10^2$	$\gamma = 10^4$	$\gamma = 10^6$

The following pages contain the actual printout of the results. The rows of the result matrix correspond to distinct values of p and the columns to distinct values of r. Both variables vary over the set

.2, .6, .8, .9, .95

Q= 0.99999E-02 ALPHA= 0.99999E-01 GAMMA= 0.00000E 00

EFFECTIVENESS TABLE

0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00

Q= 0.99999E-02 ALPHA= 0.99999E-01 GAMMA= 0.10000E 01

EFFECTIVENESS TABLE

0.00	0.01	0.01	0.01	0.01
0.00	0.01	0.01	0.01	0.01
0.00	0.01	0.01	0.01	0.01
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00

Q= 0.99999E-02 ALPHA= 0.99999E-01 GAMMA= 0.10000E 03

EFFECTIVENESS TABLE

0.19	0.59	0.79	0.89	0.94
0.19	0.59	0.79	0.89	0.94
0.19	0.59	0.79	0.89	0.94
0.19	0.59	0.78	0.88	0.93
0.18	0.57	0.76	0.85	0.90

Q= 0.99999E-02 ALPHA= 0.99999E-01 GAMMA= 0.10000E 05

EFFECTIVENESS TABLE

-0.00	-0.00	-0.00	-0.00	-0.00
-0.00	-0.00	-0.00	-0.00	-0.00
-0.00	-0.00	-0.00	-0.00	-0.00
-0.00	-0.00	-0.00	-0.00	-0.00
-0.00	-0.00	-0.00	-0.00	0.00

Q= 0.99999E-02 ALPHA= 0.99999E-01 GAMMA= 0.10000E 07

EFFECTIVENESS TABLE

-0.00	-0.00	-0.00	-0.00	-0.00
-0.00	-0.00	-0.00	-0.00	-0.00
-0.00	-0.00	-0.00	-0.00	-0.00
-0.00	-0.00	-0.00	-0.00	-0.00
-0.00	-0.00	-0.00	-0.00	-0.00

Q= 0.99999E-03 ALPHA= 0.10000E 01 GAMMA= 0.00000E 00

EFFECTIVENESS TABLE

0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00

Q= 0.99999E-03 ALPHA= 0.10000E 01 GAMMA= 0.10000E 01

EFFECTIVENESS TABLE

0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00

Q= 0.99999E-03 ALPHA= 0.10000E 01 GAMMA= 0.10000E 03

EFFECTIVENESS TABLE

0.02	0.06	0.08	0.09	0.10
0.02	0.06	0.08	0.09	0.10
0.02	0.06	0.08	0.09	0.10
0.02	0.06	0.08	0.09	0.10
0.02	0.06	0.08	0.09	0.09

Q= 0.99999E-03 ALPHA= 0.10000E 01 GAMMA= 0.10000E 05

EFFECTIVENESS TABLE

-0.00	-0.00	-0.00	-0.00	0.50
-0.00	-0.00	-0.00	-0.00	0.50
-0.00	-0.00	-0.00	-0.00	0.50
-0.00	-0.00	-0.00	-0.00	0.50
-0.00	-0.00	-0.00	-0.00	0.50

Q= 0.99999E-03 ALPHA= 0.10000E 01 GAMMA= 0.10000E 07

EFFECTIVENESS TABLE

-0.00	-0.00	-0.00	-0.00	-0.00
-0.00	-0.00	-0.00	-0.00	-0.00
-0.00	-0.00	-0.00	-0.00	-0.00
-0.00	-0.00	0.00	0.00	0.00
-0.00	-0.00	-0.00	-0.00	-0.00

Q= 0.99999E-05 ALPHA= 0.10000E 03 GAMMA= 0.00000E 00

EFFECTIVENESS TABLE

0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00

Q= 0.99999E-05 ALPHA= 0.10000E 03 GAMMA= 0.10000E 01

EFFECTIVENESS TABLE

0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00

Q=   0.99999E-05   ALPHA=   0.10000E 03   GAMMA=   0.10000E 03

EFFECTIVENESS TABLE

0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00

Q= 0.99999E-05 ALPHA= 0.10000E 03 GAMMA= 0.10000E 05

EFFECTIVENESS TABLE

0.02	0.06	0.08	0.09	0.10
0.02	0.06	0.08	0.09	0.10
0.02	0.06	0.08	0.09	0.10
0.02	0.06	0.08	0.09	0.10
0.02	0.06	0.08	0.09	0.10

Q= 0.99999E-05 ALPHA= 0.10000E 03 GAMMA= 0.10000E 07

EFFECTIVENESS TABLE

-0.00	-0.00	-0.00	-0.00	0.50
-0.00	-0.00	0.00	0.00	0.50
-0.00	-0.00	-0.00	-0.00	0.50
-0.00	-0.00	-0.00	-0.00	0.50
-0.00	-0.00	-0.00	-0.00	0.50

Q= 0.99999E-06 ALPHA= 0.10000E 04 GAMMA= 0.00000E 00

EFFECTIVENESS TABLE

0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00

Q= 0.99999E-06 ALPHA= 0.10000E 04 GAMMA= 0.10000E 01

EFFECTIVENESS TABLE

0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00

Q= 0.99999E-06 ALPHA= 0.10000E 04 GAMMA= 0.10000E 03

EFFECTIVENESS TABLE

0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00

Q= 0.99999E-06 ALPHA= 0.10000E 04 GAMMA= 0.10000E 05

EFFECTIVENESS TABLE

0.00	0.01	0.01	0.01	0.01
0.00	0.01	0.01	0.01	0.01
0.00	0.01	0.01	0.01	0.01
0.00	0.01	0.01	0.01	0.01
0.00	0.01	0.01	0.01	0.01

Q= 0.99999E-06 ALPHA= 0.10000E 04 GAMMA= 0.10000E 07

EFFECTIVENESS TABLE

0.20	0.60	0.80	0.90	0.95
0.20	0.60	0.80	0.90	0.95
0.20	0.60	0.80	0.90	0.95
0.20	0.60	0.80	0.90	0.95
0.20	0.60	0.80	0.90	0.95

Q= 0.99999E-07 ALPHA= 0.10000E 05 GAMMA= 0.00000E 00

EFFECTIVENESS TABLE

0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00

Q= 0.99999E-07 ALPHA= 0.10000E 05 GAMMA= 0.10000E 01

EFFECTIVENESS TABLE

0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00

Q= 0.99999E-07 ALPHA= 0.10000E 05 GAMMA= 0.10000E 03

EFFECTIVENESS TABLE

0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00

Q= 0.99999E-07 ALPHA= 0.10000E 05 GAMMA= 0.10000E 05

EFFECTIVENESS TABLE

0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00

Q= 0.99999E-07 ALPHA= 0.10000E 05 GAMMA= 0.10000E 07

EFFECTIVENESS TABLE

0.02	0.06	0.08	0.09	0.10
0.02	0.06	0.08	0.09	0.10
0.02	0.06	0.08	0.09	0.10
0.02	0.06	0.08	0.09	0.10
0.02	0.06	0.08	0.09	0.10

Q= 0.10000E-07 ALPHA= 0.10000E 06 GAMMA= 0.00000E 00

EFFECTIVENESS TABLE

0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00

Q= 0.10000E-07 ALPHA= 0.10000E 06 GAMMA= 0.10000E 01

EFFECTIVENESS TABLE

0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00

Q= 0.10000E-07 ALPHA= 0.10000E 06 GAMMA= 0.10000E 03

EFFECTIVENESS TABLE

0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00

Q= 0.10000E-07 ALPHA= 0.10000E 06 GAMMA= 0.10000E 05

EFFECTIVENESS TABLE

0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00

Q= 0.10000E-07 ALPHA= 0.10000E 06 GAMMA= 0.10000E 07

EFFECTIVENESS TABLE

0.00	0.01	0.01	0.01	0.01
0.00	0.01	0.01	0.01	0.01
0.00	0.01	0.01	0.01	0.01
0.00	0.01	0.01	0.01	0.01
0.00	0.01	0.01	0.01	0.01

APPENDIX II

THE BAYESIAN EFFECTIVENESS OF INFORMATION SYSTEMS

The Bayesian Value of an Information System

The usual model of decision making under uncertainty assumes that there are certain states of nature that are relevant to our decision, certain acts that are open to us for choice, and a utility index associated with each act-state pair.

Let X^i denote the i-th state of nature, $i = 1, \ldots, N$; A^k denote the k-th act open to us, $k = 1, \ldots, L$; and u_{ki} be the utility index assigned to the act-state pair (A^k, X^i), $u_{ki} = U(A^k, X^i)$.

An information structure can be most conveniently characterized as follows:

$$
\begin{array}{c|c}
 & \begin{array}{ccc} Y^1 & \cdots & Y^M \end{array} \\
\begin{array}{c} X^1 \\ \vdots \\ X^N \end{array} &
\begin{bmatrix} q_{11} & \cdots & q_{1M} \\ \vdots & & \vdots \\ q_{N1} & \cdots & q_{NM} \end{bmatrix}
\end{array}
$$

where Y^j, $j = 1, \ldots, M$, is the j-th message transmitted to us by the information system, and $q_{ij} = P(Y^j | X^i)$ is the conditional probability of the k-th message given the fact that the true state of the nature is X^i.

A rule which assigns an act to each of the possible messages is called a decision rule. We shall denote it by $A = \alpha(Y)$.

The Bayesian decision rule implies the following assumptions: (1) There is a certain prior probability associated with each state of nature; we shall denote it by $P(X^i)$, $i = 1, \ldots, N$. (2) For each message observed, a posterior probability distribution over the states of nature can be derived by using Bayes theorem. Let $P(X^i | Y^j)$ denote the posterior probability of X^i given the fact that Y^j has been observed. Then,

$$P(X^i | Y^j) = P(X^i) \cdot P(Y^j | X^i) \Big/ \sum_r P(X^r) \cdot P(Y^j | X^r);$$

(3)Let $V(A^k | Y^j) = \sum_i P(X^i | Y^j) u_{ki}$ be the expected value of A^k given the fact that Y^j has been observed. Then the Bayesian decision rule says that for each message Y^j, one should select the act $A = \hat{\alpha}(Y^j)$ such that

$$V[\hat{\alpha}(Y^j) | Y^j] = \max_k V(A^k | Y^j).$$

Let

$$P(Y^j) = \sum_r P(X^r) \cdot P(Y^j | X^r)$$

be the probability of observing the j-th message given the prior probability distribution over X and the information system χ. Then the Bayesian value of χ is

$$\hat{V}(\chi) = \sum_{j=1}^{M} P(Y^j) \sum_{i=1}^{N} P(X^i | Y^j)\, u[\hat{\alpha}(Y^j), X^i]. \qquad \text{(II-1)}$$

The Bayesian Effectiveness

As defined in the previous section, the Bayesian decision rule selects the act $A = \hat{\alpha}(Y^j)$, such that

$$V[\hat{\alpha}(Y^j) | Y^j] = \max_k V(A^k | Y^j)$$

i. e., such that

$$\sum_{i=1}^{N} P(X^i | Y^j)\, u[\hat{\alpha}(Y^j), X^i] = \max_k \left[\sum_{i=1}^{N} P(X^i | Y^j)\, u(A^k, X^i) \right]$$

Let us introduce the following notation:

$$U = \begin{bmatrix} u_{11} & \cdot & \cdot & \cdot & u_{1N} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ u_{L1} & \cdot & \cdot & \cdot & u_{LN} \end{bmatrix}$$

where $u_{ki} = U(A^k, X^i)$

$$Q = \begin{bmatrix} q_{11} & \cdot & \cdot & \cdot & q_{1M} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ q_{N1} & \cdot & \cdot & \cdot & q_{NM} \end{bmatrix}$$

where $q_{ij} = P(Y^j | X^i)$

$$P = \begin{bmatrix} p_{11} & \cdot & \cdot & \cdot & p_{1M} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ p_{N1} & \cdot & \cdot & \cdot & p_{NM} \end{bmatrix}$$

where $p_{ij} = P(X^i | Y^j)$

$$= \frac{P(X^i) \cdot P(Y^j | X^i)}{\sum_{r=1}^{N} P(X^r) \cdot P(Y^j | X^r)}$$

$$= \frac{P(X^i) \cdot P(Y^j | X^i)}{P(Y^j)}$$

The j-th column of P, denoted by $[P]_j$,

$$\begin{bmatrix} p_{ij} \\ \cdot \\ \cdot \\ \cdot \\ p_{Nj} \end{bmatrix} = \begin{bmatrix} P(X^i \mid Y^j) \\ \cdot \\ \cdot \\ \cdot \\ P(X^N \mid Y^j) \end{bmatrix}$$

is the conditional probability distribution over the states of nature if Y^j has been observed. It is then clear that the j-th column of UP, denoted by $[UP]_j$, is the set of expected utilities associated with various acts conditional on the occurrence of Y^j.

We shall now define the operator $*$. If $[B]$ represents a column vector

$$\begin{bmatrix} b_1 \\ \cdot \\ \cdot \\ \cdot \\ b_M \end{bmatrix}, \text{ then } [B]^* = \max_i \{b_i\}.$$

Let B be a matrix

$$\begin{bmatrix} b_{11} & \cdot & \cdot & \cdot & b_{1N} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ b_{M1} & \cdot & \cdot & \cdot & b_{MN} \end{bmatrix}$$

Then $B^* = ([B]_1^*, \ldots, [B]_N^*)$, where $[B]_j$ denotes the j-th column of B.

With the aid of the operator $*$, we can define the Bayesian decision rule as $\hat{\alpha}(Y^j) = A^{\hat{k}}$ such that $V(A^{\hat{k}} \mid Y^j) = [UP]_j^*$. Then the Bayesian value of an information system $\varkappa$ is given by

$$\hat{V}(\varkappa) = \sum_{j=1}^{M} P(Y^j)\,[UP]_j^*$$

where $P(Y^j) = \sum_{r=1}^{N} P(X^r) \cdot P(Y^j \mid X^r)$.

Let $\hat{E}(\varkappa)$ be the Bayesian effectiveness associated with an information system $\varkappa$. Then,

$$\hat{E}(\varkappa) = \hat{V}(\varkappa) - \hat{V}(\varkappa^o)$$

where $\varkappa^o$ denotes the null information system.

Consider the k-th component of $[UP]_j$. It is

$$\sum_{i=1}^{N} U(A^k, X^i)\,P(X^i \mid Y^j) = \frac{1}{P(Y^j)} \sum_{i=1}^{N} U(A^k, X^i)\,P(X^i)\,P(Y^j \mid X^i).$$

Let $\bar{U} = UD$

where $D = \begin{bmatrix} P(X^1) & . & . & . & 0 \\ . & & . & & . \\ . & & & . & . \\ . & & & . & . \\ 0 & . & . & . & P(X^N) \end{bmatrix}$ is a diagonal matrix whose

diagonal elements are $P(X^1), \ldots, P(X^N)$. Then the k-th component of $[UP]_j$ is simply $1/P(Y^j)\,[\bar{U}Q]_{kj}$, where $[\bar{U}Q]_{kj}$ is the k-th element of $\bar{U}Q$ and

$$[UP]_j = \frac{1}{P(Y^j)}\,[\bar{U}Q]_j.$$

Since $(1/P(Y^j)[\bar{U}Q]_j)^* = 1/P(Y^j)[\bar{U}Q]_j^*$, it follows that

$$V(\varkappa) = \sum_{j=1}^{M} P(Y^j)[UP]_j^*$$

$$= \sum_{j=1}^{M} P(Y^j)\left(\frac{1}{P(Y^j)}[\bar{U}Q]_j\right)^*$$

$$= \sum_{j=1}^{M} [\bar{U}Q]_j^*$$

$$= (\bar{U}Q)^* \xi$$

$$= [UDQ]^* \xi$$

where ξ is a column vector with M components whose values are all equal to 1.

Let P_o be the P matrix associated with the null information system $\varkappa_o$. Since $[UP_o]_j$ is the weighted average—with the weights $\{P(X^i)\}$—of the columns of U and is independent of j, we shall denote it by $[U_o]$. Then

$$V(\varkappa_o) = \sum_{j=1}^{M} P(Y^j)[U_o]^* = [U_o]^*$$

and $$E(\varkappa) = (\bar{U}Q)^* \xi - [U_o]^* = [UDQ]^* \xi - (U\pi)^*$$
where π is the vector whose i-th component is the prior probability of X^i, $P(X^i)$.

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* Dunlap and Associates, Inc.	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP N/A

3. REPORT TITLE

Data File Size and Its Relation to the Bayesian Effectiveness of an Information System

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Interim Report - 16 February 1965 to 15 April 1965

5. AUTHOR(S) *(Last name, first name, initial)*

Gagliardi, Ugo O.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
April 1965	80	7

8a. CONTRACT OR GRANT NO. AF19(628)5057 b. PROJECT NO. 2806 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) ESD-TR-65-275 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)* None

10. AVAILABILITY/LIMITATION NOTICES

Copies available from DDC.
DDC release to CFSTI is authorized.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Decision Sciences Laboratory, Electronic Systems Division, A.F. Systems Command, USAF, L.G. Hanscom Field

13. ABSTRACT

A simple Bayesian measure of system effectiveness for information retrieval systems is proposed. The measure combines the recall and precision ratios of an information system with the utility structure of the system user. Using the measure, it is possible to show that effective systems are possible only under a very narrow set of conditions. In particular, it is shown that using present state-of-the-art indexing, it is not possible to have effective systems with file sizes much in excess of 100,000 documents.

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Decision Making Information Retrieval System Analysis Statistical Analysis Experimental Data Models						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2*a*. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2*b*. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7*a*. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7*b*. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8*a*. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8*b*, 8*c*, & 8*d*. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9*a*. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9*b*. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through
____."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through
____."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through
____."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, *or (U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.

EDITOR'S CHECKLIST FOR TECHNICAL REPORT FORMAT

ESD TDR NUMBER
62-275 (Dunlap Associates)

DATE
14 Jun 65 (Cont. AF19(628)5057

FRONT COVER

- [ ] NO DISCREPANCIES
- [ ] COVER OMITTED
- [ ] CLASSIFICATION MARKING OMITTED
- [ ] ESD IDENTIFICATION OMITTED
- [X] IMPROPER COVER STOCK
- [ ] IMPROPER FORMAT
- [ ] DATE OMITTED
- [x] OTHER Please use Sample Cover attached - Print on 110 lb commercial weight stock (Par. 18, ESD Exhibit 65-1) Omit plastic overlay on Final Copy.

SPECIAL NOTICES (Inside Front Cover)

- [ ] NO DISCREPANCIES
- [ ] ALL SPECIAL NOTICES OMITTED
- [ ] AVAILABILITY NOTICE OMITTED
- [ ] DISSEMINATION NOTICE OMITTED
- [ ] DISPOSITION NOTICE OMITTED
- [ ] LEGAL NOTICE OMITTED
- [x] OTHER Special Notices are now included on reverse of Front Cover.

TITLE PAGE (First right-hand page)

- [X] NO DISCREPANCIES
- [ ] AUTOMATIC DOWNGRADING NOTICE OMITTED
- [ ] ESPIONAGE NOTICE OMITTED
- [ ] CLASSIFICATION NOTICE OMITTED
- [ ] TITLE PAGE OMITTED
- [ ] TITLE OMITTED
- [ ] OTHER

FOREWORD (Reverse of Title page)

- [x] NO DISCREPANCIES (Included by Decision Sciences)
- [ ] FOREWORD PAGE OMITTED
- [ ] SYSTEM, PROJECT OR TASK NUMBER OMITTED
- [ ] CONTRACTOR'S NAME AND ADDRESS OMITTED
- [ ] DISTRIBUTION LIMITATION REASON OMITTED
- [ ] CONTRACT NUMBER OMITTED
- [ ] DATE OMITTED
- [ ] CAPTION OMITTED
- [x] OTHER Foreword should be printed on reverse of Title Page. (Ref. par. 23, ESD Exhibit 65-1)

APPROVAL STATEMENT (Last para of Foreword)

- [X] NO DISCREPANCIES (Included by Decision Sciences)
- [ ] APPROVAL STATEMENT OMITTED
- [ ] OTHER

ABSTRACT (Next right-hand page)

- [x] NO DISCREPANCIES (Included by Decision Sciences)
- [ ] ABSTRACT OMITTED
- [ ] EXCESSIVE LENGTH
- [ ] OTHER

(Continued on reverse side)

TEXT

- [ ] NO DISCREPANCIES
- [ ] NOT DIVIDED INTO SECTIONS
- [ ] CLASSIFICATION MARKING OMITTED
- * [x] PAGE NUMBERING OMITTED on appendices.
- [ ] PAGE NUMBERS IMPROPERLY PLACED
- [x] PAGES PRINTED ONE SIDE ONLY (Par. 37, Exhibit)

*Page No. on supplementary matter are a continuation of Text Page Nos. (Ref. par. 30, ESD Exhibit 65-1)

- [ ] DOUBLE SPACING
- [ ] IMPROPER FOLDOUTS
- [x] OTHER

See Item #1, below.

DD FORM 1473 (Last printed page)

- [ ] NO DISCREPANCIES
- [x] FORM 1473 OMITTED (Ref. par. 34, ESD Exhibit 65-1)
- [ ] INCOMPLETE

See Item #2, below.

- [ ] OTHER

BACK COVER

- [ ] NO DISCREPANCIES
- [ ] COVER OMITTED
- [ ] CLASSIFICATION MARKINGS OMITTED

- [x] OTHER

Improper cover stock, 110 lb required (Par. 18, 65-1)

BINDING

- [ ] NO DISCREPANCIES
- [x] IMPROPERLY BOUND Reports should be stapled and bound - spiral or comb. binding will not be used. (Ref. par. 47, ESD Exhibit 65-1.)

- [ ] OTHER

ADDITIONAL REMARKS

1. Suitability for Microfilming. ESD Technical Reports are subject to further reproduction by microfilming. Final copy should be black characters on white opaque paper, (60 lb commercial weight) printed on both sides, single spaced with double spacing between paragraphs. It is realized this is a draft copy and is acceptable as such. (Ref. par. 37 and 45 ESD Exhibit 65-1.)

2. I have included a partially completed DD Form 1473; Item 7a should be completed when total page count has been determined. The key words are included on the reverse side, Roles and Links may be omitted. One completed copy of DD Form 1473 must be included as the last printed page in the report. (Ref. par. 34, Exhibit 65-1.)

Thomas L. Wreck

THOMAS L. WRECK AF31434358
TSgt, NCOIC Document Preparation Branch
Scientific and Technical Information Division
L. G. Hanscom Field, Bedford, Mass. 01731
274-6100, Ext. 4535/4539